View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Lehigh University: Lehigh Preserve

Lehigh University
Lehigh Preserve

Theses and Dissertations

1993

Graphical modeling of electron transfer coupling pathways between proteins

Jonggu Moon
Lehigh University

Follow this and additional works at: http://preserve.lehigh.edu/etd

Recommended Citation
Moon, Jonggu, "Graphical modeling of electron transfer coupling pathways between proteins" (1993). *Theses and Dissertations.* Paper 230.

This Thesis is brought to you for free and open access by Lehigh Preserve. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Lehigh Preserve. For more information, please contact preserve@lehigh.edu.

AUTHOR:

Moon, Jonggu

TITLE:

"Graphical Modeling of Electron Transfer Coupling Pathways Between Proteins"

DATE: January 16, 1994

"Graphical Modeling of Electron Transfer Coupling Pathways Between Proteins"

by

Jonggu Moon

A Thesis

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Computer Science

Lehigh University

December 10, 1993

This thesis is accepted and approved in partial fulfillment of the requirements for the Master of Science

December 10, 1993
Date

Dr. Robert F. Barnes
Lehigh University
Thesis Advisor

Dr. David N. Beratan
University of Pittsburgh
Co-Advisor

Chairperson of Department

Acknowledgments

The author wishes to thank Dr. Barnes for his assistance and guidance in bringing this thesis project to completion including the ironing out of issues involved in doing a project outside of Lehigh University. The author also wishes to thank Dr. Beratan for providing the ideas and support that made this thesis project possible. The author also is especially indebted to Linda Jacoby for her skillful assistance in overcoming a variety of administrative issues that arose. The author also thanks Dr. Steven Risser who as a post-doctoral researcher in Dr. Beratan's group supplied the subroutines that produced the coupling values that are used in this program.

This work was performed in Dr. Beratan's research group at the department of chemistry at the University of Pittsburgh, Pittsburgh Pennsylvania. The Inter-Protein Coupling software and source codes are available from Dr. Beratan. A copy of the source code is on file at the Department of Electrical Engineering and Computer Science at Lehigh University Bethlehem,Pennsylvania.

Table of Contents

List of Figures

Abstract

Electron transfer tunneling between proteins is a key component in such important biological activities as respiration and photosynthesis. The pathways of electronic coupling between donor and acceptor sites heavily influence the rate of electron transfer that may occur. A program named "Inter-Protein Coupling" written in FORTRAN for Silicon Graphics (tm) workstations will display two proteins on the screen in 3-D. The user may rotate either protein and modify the distance between them. The program shows interactively the resulting effects on the cloud of pathway coupling interactions between proteins.

INTRODUCTION

Electron transfer reactions are ubiquitous in chemistry, biology and physics. Computer graphics and graphing theory can assist the understanding of electron transfer coupling by allowing the visualization of its origins and effects that.

Electron transfer reactions in large molecules involve long distance electron tunneling between localized sites. The rate of electron transfer depends on the strength of the coupling mediated by the molecules and is a function of their chemical composition, bonding and geometry.

The goal of this project is to develop a graphical tool that can assist chemists in visualizing the interactions between electron donor and acceptor groups as two proteins are docked in any given geometry.

chapter 1

An Overview of Electron Transfer Between Proteins

Proteins

Proteins are folded linear heteropolymers of the 20 naturally occurring amino acids*(1)*. Each protein has a unique amino-acid sequence that due to the properties of each amino-acid link in the chain, results in a unique 3D structure for the overall protein.*(2)* The unique shapes are related to the bio-chemical function of the proteins. Typical Electron Transfer proteins are 35 Angstroms in diameter.

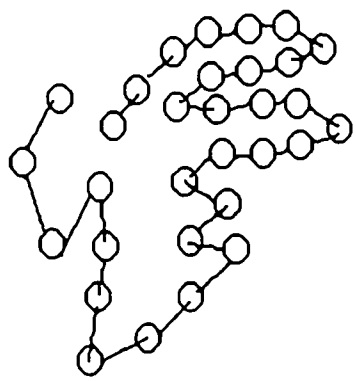

Figure 1.1: A typical protein as a chain of amino acids giving a unique shape. Each ball represents one amino acid.

Electron tunneling

In classical Newtonian mechanics, an electron would be restricted to near the atomic nuclei. Electrons are quantum particles however, so they also exhibit properties of a wave that allows it to pass through classical barriers. This effect is called "tunneling".*(3)*

As a result, electrons that are not normally expected to leave one region of one protein and be transferred to another protein, manage to do so with a non-zero probability. Many biological reactions take advantage of these changes in electron positions and the resulting changes in energies that they represent.

Coupling Pathways

As a wave like object, it is not appropriate to define a specific path that the electron will take as it is transferred from one protein to the other. Instead, the transfer is referred to as being facilitated by "coupling" between the donor and acceptor group. However, the presence of mediating groups in-between the donor and acceptor sites can enhance or diminish the rate of electron transfer.

(See figure 1.2)

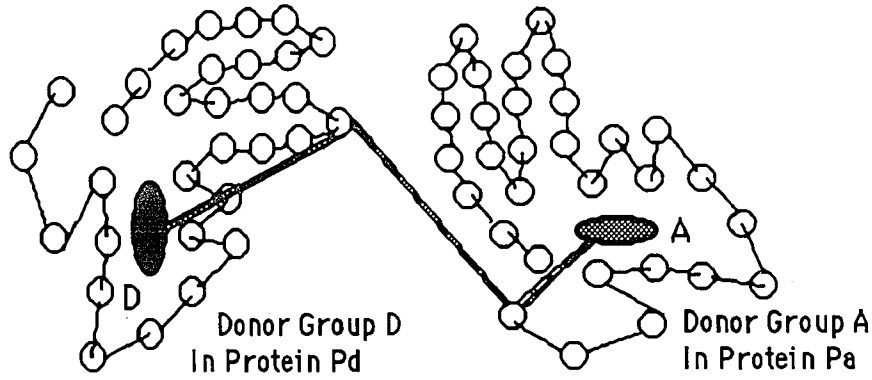

Figure 1.2: Example Coupling Pathway

It then can be said one "pathway" made up of a specific set of mediating sites is stronger than another (See figure 1.3). The strengths of these pathways are proportional to the strengths of individual chemical contacts in the protein. *(4)*

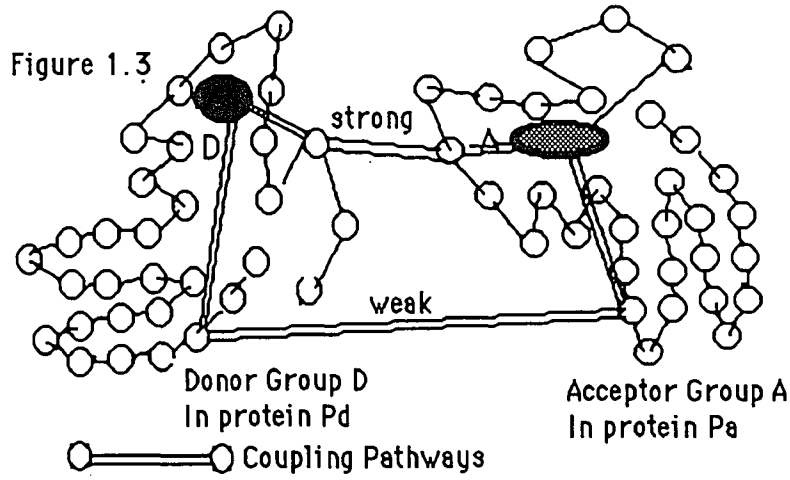

Figure 1.3 Example Comparative Coupling Pathways

Geometry affects pathways

Given that the distance protein structure and geometry strongly effect the pathway strength, altering the orientation of two proteins with respect to each other alters the pathways that best mediate the electron transfer coupling.

The geometry of the proteins can be altered by either changing the distance between proteins or their orientation rotation angles theta Θ or phi Φ.

Purpose

The purpose of this project is to develop a software tool that will display any two proteins and allow the user to orient them. Interactively. The program will then display significant coupling pathways as a "cloud of interactions" between the proteins.

Chapter 2

FEATURES

The following features are available in this version of Inter Protein Coupling.

PDB Compatibility

This software tool will accept the standard crystalographic data format known as the "Brookhaven Protein Data Bank" or PDB format. There is PDB data available for protein structures that have known xray crystal structures in ASCII files. *(4)* Since the choice of geometric center and initial state of rotation for any protein described in the PDB file is somewhat arbitrary, the Inter Protein Coupling software can only display rotation angles that are, at best, relative to the initial state of the protein in the data file. It is not relative to some standardized orientation.

At Startup

When initiated, Inter Protein Coupling will read in four files: the pathways data and PDB format data for protein 1 and the pathways data and PDB format data for protein 2. The pathways data files will be supplied by a software package called "Pathways" *(5)*

BiMolecular Display

The Proteins

Protein 1 will be displayed on the left side while protein 2 will be displayed on the right. The proteins will be scaled at the beginning of the program to fit on the screen. Afterwards the screen coordinate dimensions will remain fixed. Thus any sized protein, whether it be three amino acids long or hundreds, will take up the space of approximately one quarter of the display window. Though the coupling pathways shown will be color coded according to their strength of coupling, the colors of the proteins do not represent any value at all, but are there to contrast the proteins against the background and themselves and for the needs of the user.

The Electron Transfer Coupling Pathway

In between the two proteins will be drawn several lines between any two amino acids that take part in a significant coupling pathway. For each amino-acid in a protein, the pathways data file has supplied a coupling value between that amino-acid and an active donor or acceptor group within the protein. For every amino-acid in one protein, an interprotein coupling value is calculated with every amino-acid in the other protein. For now, a simple subroutine was written by Dr. Steven Risser to take care of this feature. For some geometries, there will be enough lines drawn that they will form what appears to be a "cloud of interaction" between the two proteins.

Simplifying the Screen

Given that the quantum mechanical tunneling of electrons between two active groups will always be some value greater than zero, no matter how small, it is counter productive to display on the screen all non zero coupling values. There would be too much information on the screen to be of use to a researcher who would be interested in only the strongest pathways for any given orientation. Two methods of displaying and labeling data are required to allow this software tool to be of help to the user.

Limiting the Number of Coupling Pathways

For this program, an algorithm first calculates all the coupling strengths for a given geometry. A second algorithm will then search through the list of pathways found and remove any coupling pathway that has a strength below a "cut off value". For now, the cut off value is one tenth the value of the maximum coupling value found for that geometry.

Color Coding the Remaining Coupling Pathways

Of the list of pathways that remain, the maximum valued one is colored in red. The minimum valued ones are colored blue and the rest are scaled appropriately with green representing the middle value. A Legend is drawn at the top of the screen to display the corresponding color to the values of the electron transfer coupling pathways.

THE CONTROL PANEL

For the rest of this chapter, refer to figure 2.1 Control Panel unless otherwise indicated.

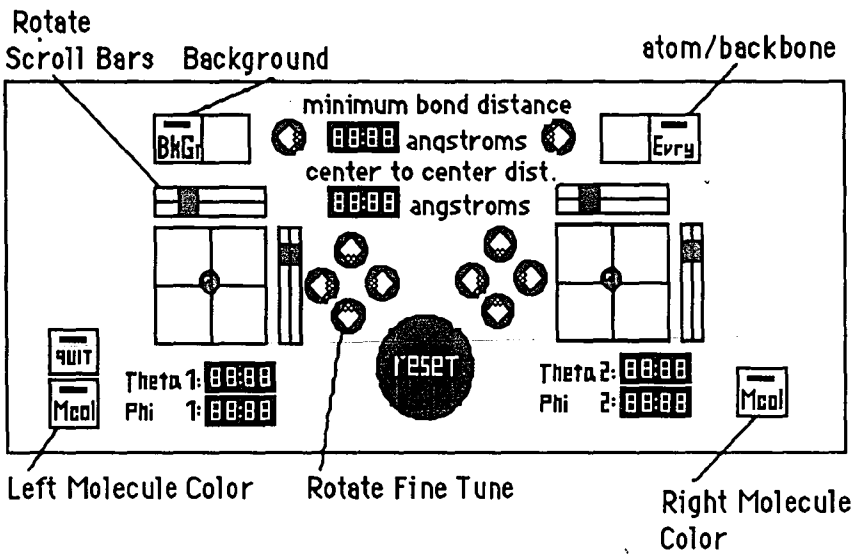

Figure 2.1 Control Panel

"Mouse Box" (joystick box) & Fine Tuning

The user is allowed to alter the relative positions of the proteins with respect to each other by the manipulation of mouse activated control devices.

Two sets of controls are available, one for each molecule. They are all activated by moving the mouse arrow within the control device image and holding down the left mouse button. The area activated will then become brighter to show its status and indicators inside that area will change to show any values being modified.

The final theta and phi values are displayed in the indicators underneath the joystick box. The values are in radians. If either value exceeds 3.14, it will be reset to -3.14 and if either value becomes lower than -3.14 it will be reset to 3.14.

Coarse Control - Joystick :

To rotate the molecule into a general position that approximates the user's interests, or to casually "browse" the possible orientations, a gold bead has been drawn in the center of a large gray box. When the user activates the joystick box area, the gold bead will move to where ever the point of the mouse arrow is indicating. This device works the same as a regular game joystick. The center of the joystick box is designated theta=0.0 and phi=0.0. Moving the bead vertically will alter the phi value and moving the bead horizontally will alter the theta value. The molecule on the screen will continue to rotate in the direction indicated on the joystick box until the left mouse button is released. The position on the joystick box is a relative position to the rate the molecule is rotated at that time. Holding the gold bead close to the center of the box will cause the molecule to rotate slowly. The further away from the center the bead is moved, the higher the rate of rotation the molecule will undergo.

Rotator Scroll Bars:

For better control over the rotations of the molecules, a vertical and horizontal scroll bar is available for each molecule.

Activating these scrolls bars with the mouse will cause the gold scroll block to move to where the mouse arrow is pointing. These scroll bars represent absolute rotations of the molecule. The extreme left or extreme bottom of the scroll bar represents a -180.0 degree rotation from the original position as described in the PDB file. The extreme right or extreme top of the scroll bar represents a 180.0 degree rotation. Holding the scroll block in the center will cause the molecule to regain the original rotation, theta and phi equal to zero.

Fine Tune Push Buttons:

The theta and phi displays are accurate to 2 significant figures after the decimal point. It has been decided that a 0.005 radian loss in the accuracy of rotation of a protein will not result in a significant change in coupling values obtained. When the Push Button is activated, whatever value is being affected will be incremented or decremented by 0.01. radians.

Protein to Protein Distance in Angstroms

The distance between the two proteins may be modified by the two Push Buttons on the control panel labeled "minimum bond distance". This is the closest that the two molecules will be allowed to approach each other. When activated, the distance between the right most component of the left molecule, and the left most component of the right molecule will be set at that distance in angstroms. The proximity of the two proteins will be figured accordingly. For convenience, a distance in angstroms between the geometric center of the two proteins are also displayed. This value is the x coordinate value of the amino acids plus the minimum bond distance.

Molecule Color Scroll Bars:

Who ever will use this program, perhaps a theoretical chemist, may have reasons to want one or both molecules to display certain colors on the screen. This may be in order to remain consistent with color coding schemes in other diagrams. To this end, the user may activate either "LED Button" labeled "Mcol" on the appropriate half of the control panel. This will cause a new panel to appear. (see figure 2.2) This new panel contains three scroll bars corresponding to the colors red, green and blue , allowing the user to vary the proportions of each color to obtain the desired combination.

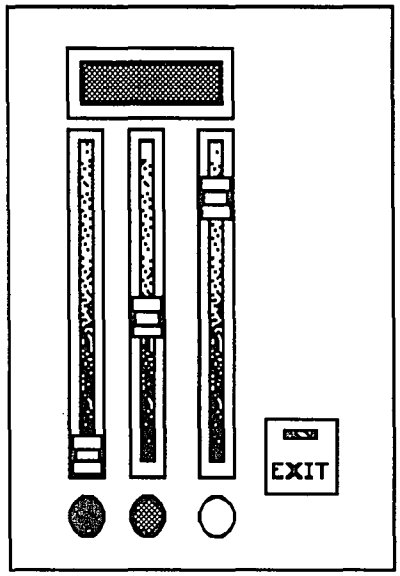

Figure 2.2
Molecule
Color Panel

Figure 2.2: Molecule Color Panel

The user can change the color values by activating the scroll blocks. The resulting color value will be displayed in the bottom circular window for each color, and an overall color will be displayed in the final color window at the top.

Everyatom / Backbone

A typical protein, such as Cytochrome C, has 103 amino-acids in a chain. Calculating the electron coupling matrix between two cytochrome C proteins will involve a matrix of size 103 x 103 or 10,609 calculations for each incremental rotation. This slows down most of today's workstations to about a half second to one second per incremental rotation. If the Inter Protein Coupling were to allow the display of all atoms in each protein, about 1,500 atoms in cytochrome C, the resulting coupling matrix will cause approximately 5 to 15 second delays per rotation increment.

Most users cannot afford the time to browse the geometries at that rate. The control panel contains a toggle-LED-button pair to allow the user to switch back and forth from displaying only the protein backbones (see figure 1.1) during rotation, and Every atom display when the desired geometry is obtained.

Black Background / White Background

It appears that in an aesthetic sense, a black background works best when the program is run on the workstation screen. However, the black background is not well suited for printing on a transparency. Instead a white background results in a clearer image on the transparency. This program supplies a toggle-LED-button pair on the upper left control panel to alternate between the two.

Reset Button

In the case that the user is unsatisfied with the new orientation of the proteins and wishes to return to the original geometry as stored in the input data files, a "reset" Push Button is supplied at the bottom of the control panel. Activating this red Push Button and holding down the left mouse button for five seconds will cause all variables to be reset back to their original values when Inter Protein Coupling was initiated.

Issues in Software Engineering

Programming Environment

The software was developed using Silicon Graphics FORTRAN under the IRIX (UNIX) operating system on a Silicon Graphics Personal IRIS. The software will run on any Silicon Graphics IRIS or INDIGO. Silicon Graphics is a trade mark of Silicon Graphics Inc. Silicon Graphics enjoys a reputation as being one of the best machines for the application of graphics and modeling software in scientific research.

Inter Protein Coupling comprises of 7 files of source code. It requires about 4.1 Mb of storage. To compile it there must also be available the include files fgl.h, fdevice.h, fget.h and fsphere.h. A general listing of the subroutines follows (figure 3.1)

IPC.F	INIT.F	CONTROLS.F	SUBS.F
main loop	init	controls	calcsurf
calcnewpos	initvals	exitsys	calcr
	readData	setEvery	calccoupl
ROTATORS.F	scaleData	setbkgr	
rotators	readpdb	setMcolor	UTILS.F
trirot		drawMCbars	colorize
tribut	DISMOL.F	cbar	depthcue
drawarrow	drawcoupl	scrblk2	LEDButton
scrollbar	drawmolecule	disvalues	WindowFrame
scrblock	drawline	legend	GreyBox
joystick		reset	ColorBox
drawbead			DrawButton

figure3.1 Listing of Subroutines

Language

C and C++ are making inroads into scientific research and more students in chemistry and physics are turning toward these and other newer programming languages. The majority of the theoretical chemists, however, still prefer FORTRAN for their programming needs. FORTRAN is still better adapted for the manipulation and calculation of large amounts of numbers.

For a typical protein that may have hundreds of amino-acids, the calculation of all possible inter protein electron transfer couplings and the application of rotation functions to the position vectors contribute the most to this software tool's time delays. Though FORTRAN is not ideal for high quality graphical displays, its number crunching ability make up for what it lacks in other areas. Also, the new FORTRAN compilers that allow the nesting of code into visual blocks that correspond to logical groupings allow the programmer to apply the principles of software engineering as readily as if the code was in Pascal or C.

This software will not remain in its present state, but will be modified and revised and grow as the users receive copies and report back with proposed changes. It will be easier on all parties concerned if Inter Protein Coupling were written in a language that the future owners of this program understand so they could modify the code themselves. For the present, that language is FORTRAN.

Graphics Library

The Graphics Library (GL) is a trademark of Silicon Graphics Inc. GL is a library of subroutines for 2 dimensional and 3 dimensional color graphics and animation. It was invaluable to this project by taking care of the actual pixel by pixel displays on the screen. They allow the software engineer to access basic routines that draw lines, polygons and circles. Other routines allow easy filling in and coloring of these shapes. A third set of routines allow easy translation and rotation of these shapes. Taken together, these library routines made it possible to develop the elaborate graphical animation presented in this thesis project in minimal time and effort.

GL allows the programmer to define in the program's memory a three dimensional space of any size. This internal world centered at coordinate 0.0, 0.0, 0.0 is the reference point from which all the drawing and viewing routines will be based. To draw a shape (line, polygon or circle) the programmer states a list of vertices for each component of that shape. (figure 3.2)

```
call bgnpol()     ; the beginning of a shape
call rot('x',90)  ; rotate the following shape 90 degrees in the x axis
call trans(x1,y1,z1) ;  translate the following shape in the
                                given x,y,z direction
call color( scolor ) ;   scolor is an array of three numbers for R,G,B.
call v3i( vert1 ) ;  a list of verticies of the shape
call v3i( vert2 ) ;   each vert* is an array of three numbers
call v3i( vert3 ) ;  to represent the x,y and z components
call v3i( vert4 ) ;  of the vector.
call endpol()     ;  The definition of the shape is complete.
```

figure 3.2, a sample shape definition

When the endpol() call is made, this shape that has been defined will be drawn with color = scolor at position x1,y1,z1 and rotated 90 degrees in the 'x' axis. GL will take care of all other issues including lighting, shading and zbuffering.

Zbuffering

When an object is drawn in 3-dimensional space, it is important for the sense of realism that it partially block an object that it is in front of, and that it be partially be blocked by any objects that are in front of it. GL handles this question by a video buffer called "Zbuffer". In GL, the X axis is the dimension that extends horizontally across the screen, the Y axis is the dimension that extends vertically up and down the screen and the Z axis is the dimension that extends into and out of the screen. (see figure 3.3)

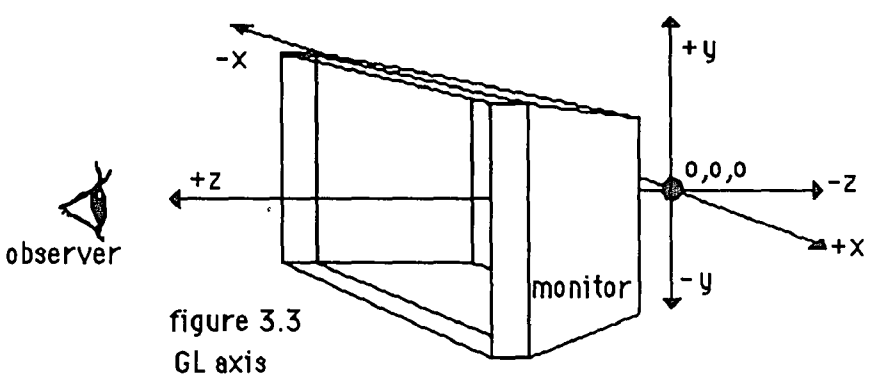

figure 3.3
GL axis

GL treats the monitor as a directional viewer with a position in the GL three space. In figure 3.3, the monitor is at position x=0, y=0, z=some_positive_value and has a direction x=0, y=0, z=negative.

As an object is being "drawn" into the GL three space, a corresponding image is being drawn into the video buffer representing how it would appear if an observer were where the monitor was, facing in the same direction as the monitor. For each pixel drawn into the video buffer, a Z value for that pixel is stored in the Zbuffer. If that Z value is closer to the viewer's Z value than the one already in the Zbuffer, that means the new pixel is closer than the older one and will block out the view of the older one. Similarly, if a new object is drawn closer that the present one, the present object will be blocked out of view according to the Z value in the Zbuffer.

Top Down Drawing Tools

Given the line, polygon and circle routines described above, more elaborate drawing routines are made possible. A few have been developed as part of this project in an effort to create an "intuitively obvious" user interface for the Inter Protein Coupling program.

Four main types of user controls appear on the control panel (see figure 2.1): Push Buttons, LED Buttons, Joystick boxes and Scroll Bars.

Push Buttons are drawn as one light circle shifted up and to the left for highlight, one dark circle shifted down and to the right for shading and one medium shaded circle in the center for the button's surface. When activated, only the surface button is drawn without the highlight nor shading to give a "pushed" appearance on the control panel.

GreyBox This is typical of many of the utility routines that have been written for this project. A routine will be passed in parameters defining the position and dimensions of a box to be drawn. A generic grey box with highlighting on the top left and shading on the bottom right will then be drawn.

LED Buttons To give the effect of a "Light Emitting Diode" a gray box is drawn first. Lettering is added with appropriate colors. Then a "color box" routine, similar to the greybox routine is called to place the light-bulb-behind-a-plate effect.

Mouse activated Joystick When four greyboxes are drawn together to make one larger greybox, their highlighted and shaded sides conveniently result in a "cross hairs" effect through the middle of the box. A routine to draw a small gold sphere that imitates the x,y coordinates of the mouse arrow when in active mode will follow the mouse arrow around.

Scroll Bars This is the most elaborate of the mouse activated user input control devices. For the rotator controls, two greyboxes dimensioned so that they are long rectangles are placed together forming a groove in-between them.

A routine to draw a gold colored "scroll box" has the x and y coordinates of the mouse arrow passed in so that if activated, the scroll box will follow the mouse arrow up and down the scroll bar. For the molecule color panel, several greyboxes and color boxes called alternately create the effect of grey frames and embedded color palettes.

Depth Cue

For a very handy three dimensional effect, a routine was written to make sections of the molecules appear brighter, clearer and larger when their Z values were positive, and appear smaller and their colors closer to the background color when their Z values were negative.

Chapter 4

Conclusions

A software tool for visualizing the electron transfer coupling pathways between two proteins of variable geometries has been presented. To my knowledge, not much literature has been devoted to the topic of the coupling effects from variable bi-molecular orientation. The emphasis here has been on the development of a software platform from which future work may be launched with special interest being paid to the graphical display and the user interface.

A print of the screen has been included to show what the program looks like at this point. Already ideas have come up regarding how to improve and expand the program beyond this initial stage. Some of those ideas will be discussed here:

Zoom in, Zoom out: Allow the user to choose a specific area on the screen of special interest and zoom the view window in on that area.

More than Two Proteins: Make the program flexible enough to accept three or more proteins at a time and still be able to rotate and translate them independently.

X Windows Compatible: As is, Inter Protein Coupling can be run from any other SGI machine that can log into the user's own SGI machine via rlogin or telnet. A desirable option would be to write in bridging code so that the graphics routines can be operated in an X Windows environment, thereby allowing the program to be run from any X Windows capable terminal through a remote link.

Alternate Data Formats: There are other protein data formats than Brookhaven's Protein Data Bank. Effort should be made to allow Inter Protein Coupling to have access to more of the most popular formats.

Output of Runtime Data: As is, this program does not have a routine to write out the results it calculates to a file. A high priority modification for researchers would be the ability to write out and read in orientation angles and distances as well as coupling pathway locations and strengths.

I suspect that a tool program such as this will not likely ever be complete. That the general applicability of the study of two amorphous objects and the interactions resulting from their relative orientations will give rise to ever new ideas and applications in scientific research.

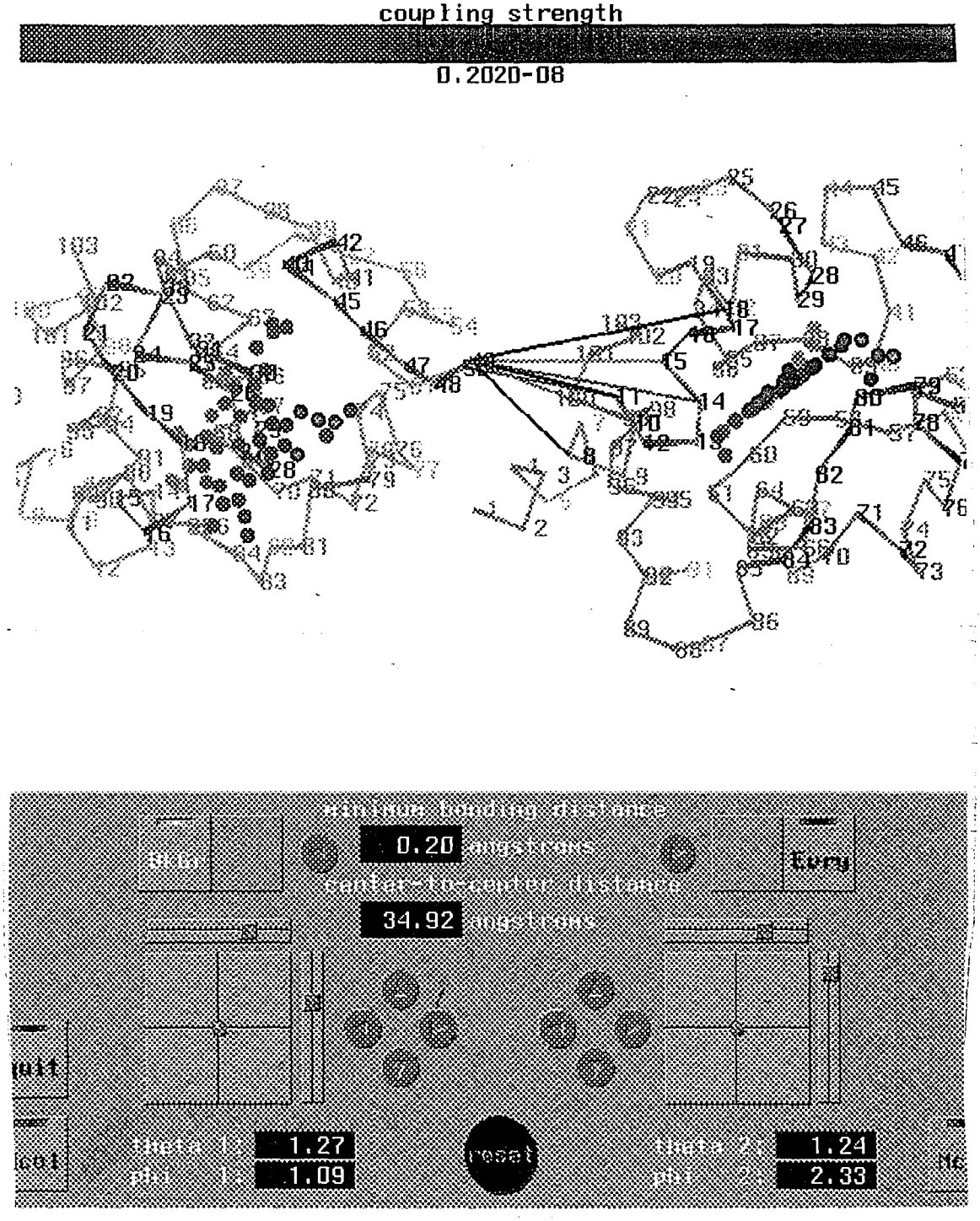

Figure 4.1: IPC Sample Image

References

1. D.D. Ebbing and M.S. Wrighton (Eds.) (1990). CHEMISTRY(3rd. ed.). Boston MA; Houghton Mifflin Company. pp 1010
2. D.D. Ebbing and M.S. Wrighton (Eds.) (1990). CHEMISTRY(3rd. ed.). Boston MA; Houghton Mifflin Company. pp 1015
3. R.A.Alberty and R.J.Silbey (1992). Physical Chemistry(1st. ed.). Canada; John Wiley & Sons, Inc. pp334
4. P.Murray-Rust Glaxo Group Research(1992). *Analysis of molecular geometry.* J.M.Goodfellow and D.S.Moss (Eds) Computer Modelling of Biomolecular Processes. London; Ellis Horwood Limited, pp21
5. J.N.Betts,D.N.Beratan,J.N.Onuchic (1992). *Mapping Electron Tunneling Pathways: An Algorithm that Finds the "Minimum Length"/Maximum Coupling Pathway between Donors and Acceptors in Proteins.,* Journal of the American Chemical Society 1992 Vol 114, pp 4043-4046.

Appendix A

Glossary of Terms

Amino-acid : 20 naturally occurring small molecules possessing an amino group and a carboxyl group.

Backbone: Representation of a protein by showing the amino-acid components as single units connected by bonds. The location of the unit is usually the location of the alpha carbon of that amino-acid.

Coupling : The interaction between a donor or donating mediating site with an acceptor or an accepting mediating site.

Depth Cue: A graphical programming effect of making objects closer to the observer seem larger and clearer.

Dock: To bring together two molecules in an optimal geometry.

Electron Transfer : The jumping of an electron from one one site to another within or between molecules.

Geometry: A orientation of two objects with respect to each other.

Mediating Site: A site that by its special properties, contribute to a coupling pathway.

Pathway: A specific series of mediating steps between donor and acceptor sites that aid in electron transfer.

PDB: Brookhaven Protein Data Bank. A data format for the representation of a protein's structure.

Protein: Folded linear heteropolymer of the 20 naturally occurring amino acids.

Tunneling: The action by which a quantum particle may pass through a barrier due to its wave like properties.

Zbuffer: One method by which a program can create a "hidden line" effect in three dimensional graphics.

Vita:

Jonggu Moon was born the only son of parents Sanggiu Moon and Choonhye Moon on March 31, 1966 in Seoul, Republic of Korea. He moved with his parents to Brooklyn, New York in 1970. They finally settled down in Nassau County, Long Island, New York where he attended John F. Kennedy High School.

He received his Bachelor of Science in Computer Science in June 1989 at Lehigh University where he participated in many extracurricular activities including the Tae Kwon Do Club, the Mountain Under the Star Medieval Club, and the Lehigh Christian Fellowship. As a graduate student at Lehigh, he held several programming positions within the University until he moved to Pittsburgh, Pennsylvania where he worked for Dr. Beratan in the Department of Chemistry in the University of Pittsburgh for his masters thesis project.

END

OF

TITLE